

Citation for published version:

Farmer, E, Jicol, C & Petrini, K 2020, 'Musicianship Enhances Perception But Not Feeling of Emotion From Others' Social Interaction Through Speech Prosody', *Music Perception*, vol. 37, no. 4, pp. 323-338.
<https://doi.org/10.1525/mp.2020.37.4.323>

DOI:

[10.1525/mp.2020.37.4.323](https://doi.org/10.1525/mp.2020.37.4.323)

Publication date:

2020

Document Version

Peer reviewed version

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Musicianship enhances perception but not feeling of emotion from others' social interaction through speech prosody

Eliot Farmer¹, Crescent Jicol¹, Karin Petrini¹

¹ Department of Psychology, University of Bath, Bath, UK

Corresponding Author:

Dr Karin Petrini

Senior Lecturer/Associate Professor

Department of Psychology

University of Bath

Claverton Down

Bath

BA2 7AY

United Kingdom

Email: K.Petrini@bath.ac.uk

Running head: Musicianship enhances perceived emotions

Abstract

Music expertise has been shown to enhance emotion recognition from speech prosody. Yet, it is currently unclear whether music training enhances the recognition of emotions through other communicative modalities such as vision and whether it enhances the feeling of such emotions. Musicians and non-musicians were presented with visual, auditory, and audiovisual clips consisting of the biological motion and speech prosody of two agents interacting. Participants judged as quickly as possible whether the expressed emotion was happiness or anger, and subsequently indicated whether they also felt the emotion they had perceived. Measures of accuracy and reaction time were collected from the emotion recognition judgements, while yes/no responses were collected as indication of felt emotions. Musicians were more accurate than non-musicians at recognising emotion in the auditory-only condition, but not in the visual or audiovisual conditions. Although music training enhanced recognition of emotion through sound, it did not affect the felt emotion. These findings indicate that emotional processing in music and language may use overlapping but also divergent resources, or that some aspects of emotional processing are less responsive to music training than others. Hence music training may be an effective rehabilitative device for interpreting others' emotion through speech.

Keywords: music expertise, emotion perception, emotional feeling, social interaction, multisensory facilitation, audiovisual processing

Introduction

Playing a musical instrument requires the development and application of numerous specialised skills, such as the rapid processing of auditory information, the translation of written notation into coordinated movements, and precise rhythmic timing. These skills develop through substantial training and practice; a complex and motivating multisensory experience that requires the integration of auditory and visual sensory information with motor responses (Edward & Carole, 1990; Lee & Noppeney, 2014; Zatorre, Chen, & Penhune, 2007). Therefore, the study of musicians provides an ideal model to explore experience-dependent changes in cognition, perception and neuroplasticity.

A central component of music training involves the production and perception of complex, expressive patterns of sound (Gardiner, 2008), and it is therefore not surprising that musicians show perceptual differences compared to non-musicians in the auditory domain. In terms of perceptual ability, musicians have been shown to be more capable of segmenting speech from background noise (Parbery-Clark, Strait, Hittner, & Kraus, 2014; however see Boebinger et al., 2015 and Madsen et al., 2019 for a lack of difference between musicians and non-musicians), and show enhanced perception of auditory features such as pitch (Kishon-Rabin, Amir, Vexler, & Zaltz, 2001; Moreno et al., 2009), timbre (Chartrand & Belin, 2006), and intensity (Hausen, Torppa, Salmela, Vainio, & Särkämö, 2013). It is plausible that these enhanced perceptual abilities extend to the perception of emotions, which, in both music and speech, are expressed through variations in acoustic features such as intonation, pitch, and intensity (Juslin & Laukka, 2003). However, the effect of music expertise on emotion processing has been sparsely studied and research has predominantly focused on emotions expressed through excerpts of musical performances.

Within the music domain, research has indicated that musicians and non-musicians do indeed perceive emotions differently. For example, length of music training is positively correlated with the accuracy with which emotions are perceived when hearing musical excerpts (i.e., when the emotion perceived by participants matches the music excerpt intended emotion; Lima & Castro, 2011a; Livingstone, Muhlberger, Brown, & Thompson, 2010), and musicians and non-musicians differ in perception of expressiveness and emotion when observing solo musical performances of a drummer (Di Mauro, Toffalini, Grassi, & Petrini, 2018). It is also possible that these differences extend to non-musical stimuli, as Juslin and Laukka (2003) have indicated that acoustic cues used to express emotions are similar in both music and speech. For example, anger in music is expressed through fast tempo, high sound level and high-frequency energy, and rising pitch. Similarly, anger in speech is expressed through fast speech rate, high voice intensity and high-frequency energy, and rising pitch (Juslin & Laukka, 2003). Thus, if common mechanisms are used for processing emotions in both domains, it is plausible that the advantage shown by musicians for processing emotion in music extends to the processing of emotion in speech.

Only a few studies have examined whether music training affects emotion perception from auditory sources beyond the music domain, although some evidence can be drawn from research on linguistic prosody; a non-verbal component of speech characterised by variations in pitch, loudness, rhythm, and timbre, that can signify the emotion of an utterance (Wu & Liang, 2011). For example, Thompson, Schellenberg, and Husain (2004) showed with three experiments that not only adult trained musicians performed better than non-musicians when recognising different emotions from prosody, but that even 6-year-old children trained with music instruments for one year were better able to recognise emotions from prosody compared to children of the same age that took drama classes for one year. Similarly, Good et

al. (2017) showed that children (aged between 6 and 15 years) who used cochlear implants and undertook six months piano lessons showed enhanced emotional speech prosody perception compared to children with cochlear implants who undertook painting lessons for six months. Finally, Lima and Castro (2011b), assessed the effect of musical experience on recognition of emotions from speech prosody, and showed that musicians perceived emotion more accurately than participants with no music training; a difference that was found across all six basic emotions (anger, disgust, fear, happiness, sadness, and surprise; Ekman, 1992). However, the stimuli consisted of single-speaker sentences, whereas studies have indicated that emotion recognition is more accurate when stimuli represent multiple people interacting with the same emotion (Cauldwell, 2000; Clarke, Bradshaw, Field, Hampson, & Rose, 2005). Furthermore, the findings of Lima and Castro (2011b) contrast with an earlier study by Trimmer and Cuddy (2008), which found no association between music training and recognition of emotional speech prosody. However, as Trimmer and Cuddy (2008) study was a correlational study, there was no cut-off specified for musicianship and thus no indication of the number of participants who were musically trained. Since recognition of emotions is a fundamental social function (Klaus, Rainer, & Harald, 2001), and therefore likely to still be efficient in non-musicians (Hauser & McDermott, 2003), it is plausible that an extensive period of music training is needed to detect differences in emotion perception (Lima & Castro, 2011b).

Although sound is a dominant feature of music, auditory information usually coincides with other elements of music practice or performance, such as translating musical notation into motor activity (Herholz & Zatorre, 2012). Furthermore, visual expressive skills are highly important to musicians (Lindström, Juslin, Bresin, & Williamon, 2003) who must learn to use body movement to visually express intention and emotion (Dahl & Friberg,

2007), and must also learn to visually interpret the intentions of other musicians in order to communicate during live performances (Gardiner, 2008). Nevertheless, few studies have explored the effects of music training on visual and multisensory perception of emotion. Vines, Krumhansl, Wanderley, Dalca, & Levitin (2011) found that when musicians were presented with the sound and video of a solo clarinet performance, the visual information augmented the emotional content perceived through the sound, compared to when the sound was presented alone. This effect did not extend to non-musicians in separate studies by Petrini, McAleer, et al. (2010) and Petrini, Crabbe, et al. (2011), who found that sound dominated the visual signal in the perception of affective expression of excerpts of drumming and saxophone performances. Additionally, Lima et al. (2016) recently showed that individuals with congenital amusia (a lifelong impairment in music perception and production) show a reduced ability to recognise emotion not only from prosody and other emotional sounds (e.g. crying) but also from silent facial expressions, indicating that impairment in music processing affects emotion recognition beyond the auditory domain. Taken together, these findings suggest that musicians may place greater weight on emotional visual information compared to non-musicians; at least when perceiving emotions from musical performances. This suggestion has been recently supported by Di Mauro et al. (2018), who found that musicians placed greater weight on visual information when perceiving emotion from drumming clips. However, the effect of music expertise on visual and multisensory perception of emotion has only been tested using musical stimuli, where musicians have more expressive knowledge, and so it remains unclear as to whether this effect extends to visual and multisensory information in non-musical domains, where both musicians and non-musicians can be considered to have a similar level of expertise.

Finally, it remains unclear whether music training affects the feeling of emotion. Indeed, it has been postulated that the perception of emotion, whether exhibited by another person or within a piece of music, may directly induce the same emotion in the perceiver (Hatfield, Cacioppo, & Rapson, 1993; Juslin, Harmat, & Eerola, 2014). For example, Neumann and Strack (2000) have shown that listeners often feel the emotions portrayed in the vocal expressions of other people, and similarly, Juslin et al. (2014) were able to induce sadness in listeners by playing music featuring a voice-like cello timbre performing a song with slow tempo, legato articulation, and a low volume; acoustic features consistent with the vocal expression of sadness (Juslin & Laukka, 2003). In both cases, the induction of emotion requires the detection of sound patterns, and therefore if musicians do show enhanced recognition of emotion, it is possible they would be more likely to feel such emotions themselves. However, the mechanisms through which emotions are induced extend beyond the mere perception of emotion and include other factors such as subjective appraisal of the stimulus (Roseman & Smith, 2001; Scherer, 1999). Consequently, emotions felt do not always coincide with emotions perceived (Juslin & Laukka, 2004; Scherer, 1999), and it is therefore unclear whether any perceptual benefit of music training would affect emotion induction.

The aim of the current study was to investigate whether music expertise affects the perception and feeling of emotions expressed by others' social interactions. We used a recently created and validated set of audiovisual clips depicting two people interacting (Piwek, Petrini, & Pollick, 2016) with angry or happy emotions, and presented the clips as auditory-only, visual-only, and audiovisual combined, to test 1) whether the previously found benefit for the combined clips in non-musicians (Piwek, Pollick, & Petrini, 2015) extends to musicians; 2) whether musicians' enhanced ability to recognise emotions from musical

gestures extends to non-musical gestures, thus indicating whether musicians' enhanced ability to recognise emotions is specific to speech prosody; 3) whether musicians' enhanced ability to recognise emotions extends to their feeling of such emotions. Specifically, we tested for differences in how accurately and quickly musicians and non-musicians recognised the emotion in each clip, and in how musicians and non-musicians felt the perceived emotions. This knowledge is necessary to further our understanding of how music training affects cognitive and emotional processes, of whether emotional processing in music and language share resources beyond those specific to sound (Lima & Castro, 2011b), and to indicate to what extent music can be used as an effective therapeutic device for individuals with emotion-processing disorders (e.g. autistic persons; Sharda et al., 2018).

In light of previous research indicating that musicians detect emotions more accurately in music and speech (Lima & Castro, 2011a; 2011b), together with the acoustic similarities between emotion expression in music and speech (Juslin & Laukka, 2003), we hypothesised that musicians would perceive emotions in the auditory-only stimuli faster and more accurately than non-musicians. Furthermore, given the greater weight musicians place on visual information in emotion perception from music performance (Di Mauro et al., 2018; Vines et al., 2011), and their improved multisensory abilities in other cognitive domains (Lee & Noppeney, 2011, 2014; Petrini et al., 2009), we hypothesised that musicians would also be able to detect emotions faster and more accurately than non-musicians in the visual-only and audiovisual conditions. Finally, because some mechanisms of emotion induction depend on recognition of affective cues (Juslin et al., 2014; Neumann & Strack, 2000), it was hypothesised that if musicians were more accurate at recognising emotion, they would also feel the emotion they had perceived more frequently than non-musicians. However, because

the induction of emotion involves mechanisms that extend beyond the perception of emotion alone (Roseman & Smith, 2001; Scherer, 1999), this latter hypothesis remained exploratory.

Method

Participants

The number of participants in each group was similar or higher than previous studies investigating the effect of music expertise on cognitive and perceptual abilities (e.g., Petrini et al., 2010a, 2011; Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011; Lee & Noppeney, 2014; Lu, Paraskevopoulos, Herholz, Kuchenbuch, & Pantev, 2014; Lima & Castro, 2011b). As estimated in a previous study (Di Mauro et al., 2018) the effect size reported in previous research examining differences in emotion perception and audiovisual perception between musicians and non-musicians is medium to very large (Cohen's $d \approx 0.50$ – 2 ; e.g., Bhatara et al., 2011; Lima & Castro, 2011a; Castro & Lima, 2014; Lee & Noppeney, 2014; Lu et al., 2014; Petrini et al., 2010a). Hence, we expected at least a medium effect size which required a total sample size of 20. We calculated this sample size by using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) and running an a priori power analysis for an ANOVA repeated measures within-between interaction and assuming a Cohen's F effect size equal to 0.25 (medium effect size), a level of power of 0.80, 2 groups, 6 measurements (2 emotions x 3 modalities), and an alpha level of 0.05.

40 participants were recruited through opportunity sampling. 20 participants (6 males) were non-musicians, with ages ranging from 18-29 years ($M = 21.70$, $SD = 2.56$). 20 participants (7 males) were musicians, with ages ranging from 21-28 years ($M = 22.25$, $SD = 2.10$). Non-musicians had no music training other than basic music classes within the school

curriculum, whereas musicians had at least 5 years of music training (ranging from 5-24 years; $M = 10.15$, $SD = 5.86$). Musicians were instrumentalists who played guitar ($n = 10$), piano ($n = 10$), violin ($n = 2$), trumpet ($n = 2$), saxophone ($n = 2$), clarinet ($n = 1$), and flute ($n = 1$). All participants were fluent English speakers, had normal hearing, and had normal (or corrected to normal) vision. The study received ethical approval from the Department of Psychology Research Ethics Committee at the University of Bath [17-262], and written informed consent was obtained from all participants.

Materials and stimuli

The stimuli were selected from a set of audiovisual affective clips developed and validated by Piwek et al. (2016). The clips represented the biological motion of two people interacting in the form of point-light displays, which offer the benefit of removing contextual information such as clothing or body appearance (Johansson, 1973). The clips also included the dialogue in form of prosody of the two people (see details below) which, in addition to the point-light displays, conferred information about movement and some morphological characteristics of the speakers (e.g., body size as depicted by formant spacing; Pisanski et al., 2014).

This helps prevent emotional bias that may be associated with certain cues such as identity, and ensures that visual attention is primarily focussed on body movement and expressivity (Hill, Jinno, & Johnston, 2003). The dialogue in each clip was either a deliberation consisting of two affirmative sentences (Actor 1: “I want to meet with John”; Actor 2: “I will speak to him tomorrow”) or an inquiry consisting of a question and answer (Actor 1: “Where have you been?”; Actor 2: “I have just met with John”).

From the set of clips in Piwek et al. (2016) we selected eight angry and eight happy audiovisual displays that portrayed the emotion at medium intensity, and where the emotion

had been identified previously with 75% accuracy (Piwek et al., 2016). These criteria were used to avoid any ceiling effects, whilst ensuring the correct emotion could still be identified above chance. These clips were then edited in Adobe Premiere Pro 2017 to produce auditory-only (where the video was replaced with a black background) and visual-only (where the audio track was muted) versions.

Multisensory facilitation is greater when the different senses have the same reliability (Ernst & Banks, 2002). As such, a high-cut filter attenuating sound above 280Hz was applied to the auditory stimuli using Adobe Premiere Pro 2017, in order to maintain the vocal prosody and intonation while decreasing the auditory reliability to a level more similar to the visual stimuli (similarity in reliability between the auditory and visual information when using this filter was pre-assessed by running a small pilot study). The high-cut filter was chosen as a means of reducing the reliability of the auditory information whilst emulating real-life conditions and therefore maintaining ecological validity (Knoll, Uther, & Costall, 2009; Scherer, 2003). Finally, the average sound amplitude of each clip was normalised to -0.5dBFS using Adobe Premiere Pro 2017 to ensure consistency in volume. The clips were exported as MPEG-4 (mp4) files with a resolution of 800 by 600 pixels, a frame rate of 30fps, and Advanced Audio Coding (AAC) audio with a sample rate of 44.1kHz.

Thus, the final set of stimuli consisted of 48 clips (each lasting between 2500 and 3500ms), comprised of 2 emotions (happiness and anger) expressed by 8 different pairs of actors and presented as 3 stimulus types (visual-only, auditory-only, and audiovisual). The stimuli were presented using a 15" MacBook Pro laptop with retina display, and Beyerdynamic DT 880 Pro headphones.

We focused our investigation on anger and happiness. These two emotions were selected because they are both basic emotions recognised in all cultures (Ekman, 1992), and are easy for actors to accurately convey in various scenarios (Ma, Paterson, & Pollick, 2006;

Pollick, Paterson, Bruderlin, & Sanford, 2001). Furthermore, anger and happiness are the most frequently reported emotions when people are asked to reflect upon their experienced emotions, and are also most commonly experienced as pure emotions, as opposed to combinations of multiple other emotions (Scherer & Tannenbaum, 1986). Finally, both share similar acoustic and visual properties such as high voice intensity and large movements (Dittrich, Troscianko, Lea, & Morgan, 1996; Juslin & Laukka, 2003), thus making the discrimination between these emotions a challenging task for both musicians and non-musicians.

Procedure

Participants were tested individually. On arrival, they were given information about the study, and asked to confirm their history of music training (to ensure they met the criteria for musician or non-musician). After giving informed consent, participants were seated approximately 55cm from the screen of the laptop, and wore the headphones with an intensity at the sound source of 60 dB.

Participants were told they would be watching clips of two people interacting, and that some clips would be composed by video and audio, while others only by video, or only by audio. Participants were instructed to make two consecutive responses for each clip. First, a forced-choice identification of the emotion expressed in the clip by pressing the “1” key on the laptop keyboard to indicate “happy”, or the “3” key to indicate “angry”. Participants were asked to respond as quickly as possible when giving this response. Following this initial response, participants were instructed to indicate whether they felt the emotion they had perceived by pressing either the “1” or “3” key on the laptop keyboard to indicate “Yes” or “No”, respectively. Participants were first presented with three randomly selected practice trials, containing one visual-only, one auditory-only, and one audiovisual clip. Following the

practice trials, participants had the opportunity to ask any questions, and were then presented with the full stimuli set of 48 clips, which were presented in a random order. Each trial began with a black screen and text saying “Test Clip”, which lasted for 1 second. This was immediately followed by the clip. Reaction times were measured from onset of the clip until the first key was pressed. The next trial began immediately after the participant made their second response. MATLAB Release 2017b (MathWorks, Inc., 2017) software with Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) was used to present the stimuli and collect the responses.

Results

The average value for each dependent measure was calculated from the 8 actor interaction clips per condition for each participant. Data from a single trial for one participant was removed from subsequent analysis because Matlab failed to display the clip, and so their 7 remaining data for this condition were used to calculate the average value. The remaining 47 clips for this participant, and all 48 clips for every other participant were presented successfully. Three separate mixed ANOVAs were used to analyse the 3 dependent measures. The assumptions of homogeneity of variance and sphericity were tested using Mauchly's test of sphericity and Levene's test for equality of variances. Both tests were non-significant ($p > .05$) for each ANOVA, indicating these two assumptions had been met. However, visual inspection of Q-Q plots indicated that the assumption of normally distributed residuals was violated for the ANOVA on proportion of emotions felt, and outliers were detected in the reaction time data. Data treatment in relation to these violations is discussed in the relevant sections below. Significant findings were followed up with pairwise comparisons, with Bonferroni correction applied to maintain control of family-wise type 1 error.

Recognition accuracy

A response was considered correct when it matched the intended expression of the interaction in the stimuli. The average proportion of correct responses were submitted to a mixed ANOVA, with “musical experience” (musician or non-musician) as a between factor, and with “emotion” (happy and angry) and “modality” (visual-only, auditory-only, and audiovisual) as within factors. There was a significant main effect for “emotion” $F(1,38) = 9.181, p = .004, \eta^2 = .195$, indicating that participants identified a higher proportion of emotions correctly when judging happy (.75) compared to angry (.63) clips. Additionally, the main effect of “emotion” was independent of “modality” or “musical experience”, as there was no significant interaction between “modality” and “emotion” $F(2,76) = 0.726, p = .487$ or between “musical experience” and “emotion” $F(1,38) < 0.001, p = .986$.

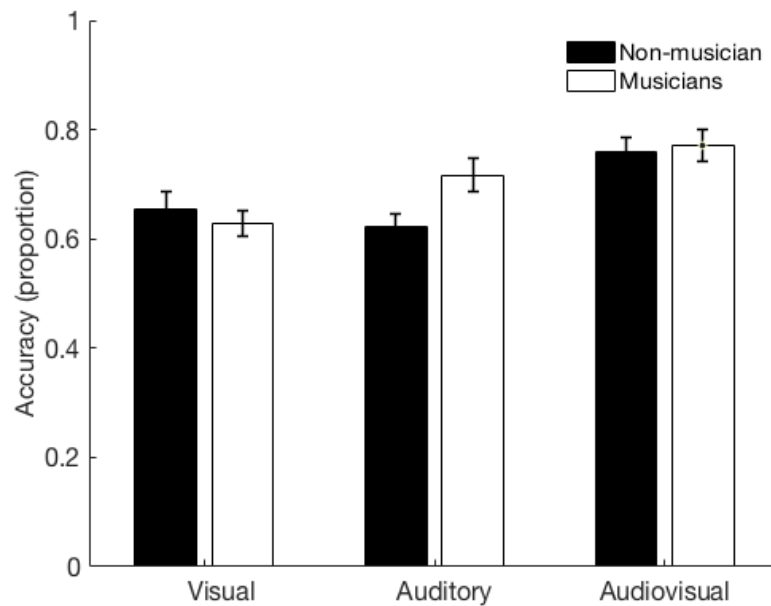


Figure 1. Proportion of correct emotion judgements given by musicians and non-musicians in the three stimuli modalities when collapsing across emotion. Error bars represent standard errors.

There was no significant main effect for “musical experience” $F(1,38) = 0.812, p = .373$, but there was a significant main effect for “modality” $F(2,76) = 15.096, p < .001, \eta^2 = .284$, and a significant interaction between “modality” and “musical experience”, $F(2,76) = 3.460, p = .036, \eta^2 = .083$ (Figure 1). Pairwise comparisons with Bonferroni correction indicated that non-musicians identified a higher proportion of emotions correctly in the audiovisual clips (.76) than in the auditory-only (.62; $p < .001$; 95% CI [0.062, 0.217]) and visual-only (.65; $p = .006$; 95% CI [0.026, 0.188]) clips, but no differences were found between the auditory-only and visual-only clips ($p = 1.000$). In contrast, musicians identified a higher proportion of emotions correctly in the audiovisual clips (.77) than in the visual-only clips (.63; $p < .001$; 95% CI [0.059, 0.221]), but no difference was found between the audiovisual and auditory-only clips (.72; $p = .288$), or the auditory-only and visual-only clips ($p = .067$). Furthermore, musicians identified a higher proportion of emotions correctly compared to non-musicians in the auditory-only clips ($p = .019$; 95% CI [-0.175, -0.016]), whereas there were no significant differences between musicians and non-musicians for the visual-only ($p = .55$) and audiovisual ($p = .83$) clips. No significant 3-way interaction between “emotion”, “modality”, and “musical experience” was found, $F(2,76) = 0.019, p = .982$. An additional analysis was conducted on the association between years of music training and recognition accuracy for all the six different conditions (2 emotions: angry and happy x 3 modalities: auditory-only, visual-only and audiovisual). Years of music training did not significantly predict the level of accuracy in the musicians’ group for any of the conditions, $F[1,18] \leq 2.035, p \geq 0.171$.

Reaction time

Average reaction times were submitted to a second mixed ANOVA with the same factors as described above. Errors and outliers (reaction times exceeding the mean of each

participant by 2 *SD*) were not included in the analysis. There was no significant main effect for “modality”, $F(2,76) = 1.660, p = .197$, or “musical experience”, $F(1,38) = 0.005, p = .946$. Concerning “emotion”, reaction times were shorter for happy clips than for angry clips $F(1,38) = 8.901, p = .005, \eta^2 = .190$, but there was a significant interaction between “emotion” and “musical experience”, $F(1,38) = 5.030, p = .031, \eta^2 = .117$ (Figure 2).

Pairwise comparisons with Bonferroni correction showed that for musicians, reaction times were significantly shorter when responding to the happy clips (1.207ms) compared to the angry clips (1.386ms; $p = .001$; 95% CI [-0.278, -0.081]). No significant differences were found between happy (1.296ms) and angry clips (1.321ms) for non-musicians ($p = .603$), or between musicians and non-musicians for either happy ($p = .615$) or angry ($p = .734$) clips.

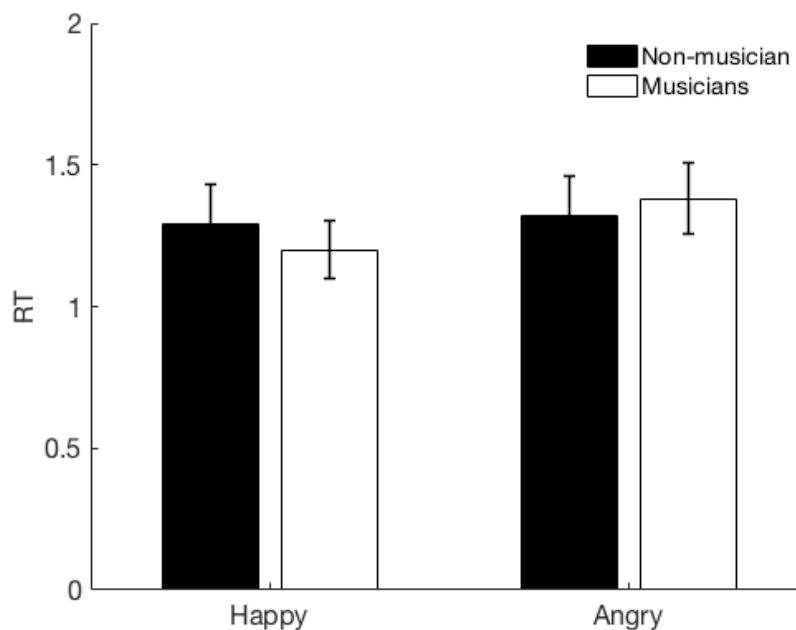


Figure 2. Mean reaction time (ms) of emotion judgements given by musicians and non-musicians in response to happy and angry clips when collapsing across stimuli modality. Error bars represent standard errors.

No significant interaction between “modality” and “musical experience” [$F(2,76) = 0.655, p = .522$], “modality” and “emotion” [$F(2,76) = 2.629, p = .079$], or “modality”, “musical experience”, and “emotion” [$F(2,76) = 2.054, p = .135$] was found.

An additional analysis was conducted on the association between years of music training and reaction time for all the six different conditions (2 emotions: angry and happy x 3 modalities: auditory-only, visual-only and audiovisual). Years of music training significantly predicted the reaction time in the musicians’ group for all conditions, $F[1,18] \geq 5.994, p \leq 0.025$ (Figure 3).

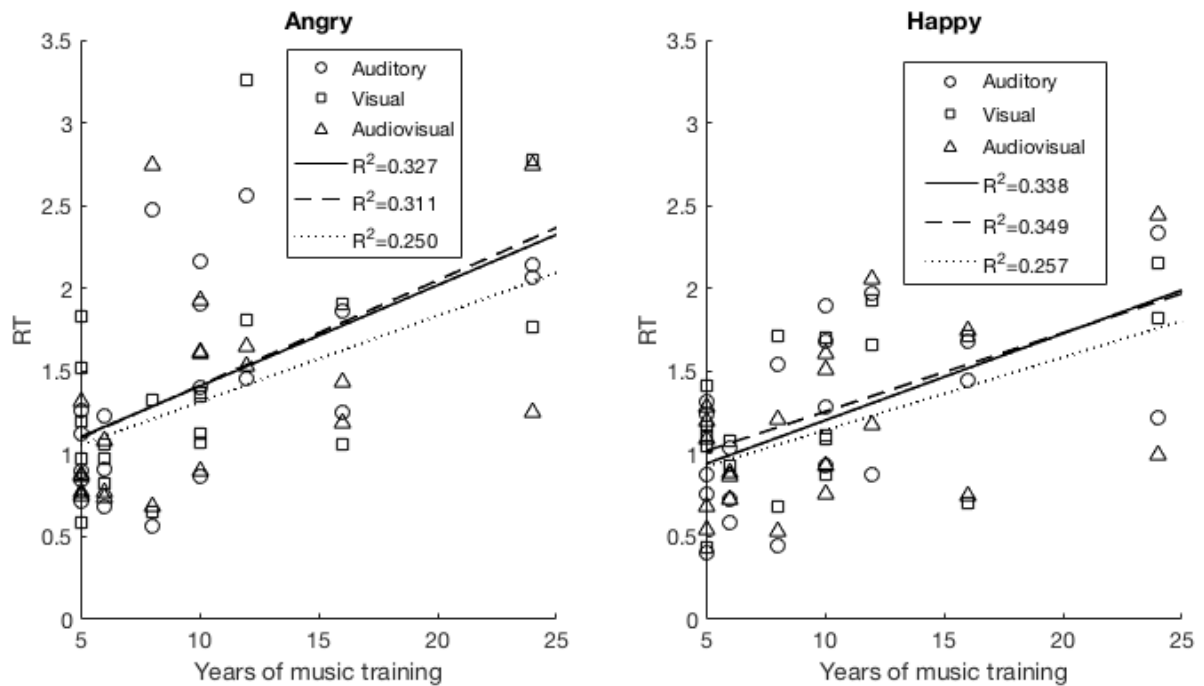


Figure 3. Mean reaction time (ms) of emotion judgements plotted as a function of years of music training.

Finally, an analysis was conducted to examine whether there was any association between accuracy responses and reaction time for musicians and non-musicians separately. We found only two significant associations for the non-musician group between the accuracy and RT

responses for happy auditory-only ($r = -.847, p < .001$) and for happy audiovisual ($r = -.512, p = .021$). No significant associations of this type were found for musicians.

The same analyses carried out on the average RT only for accurate responses returned the same results (see supplemental material).

Felt emotion

A third mixed ANOVA, with the same factors as above, was used to analyse the average proportion of responses where participants reported feeling the emotion they had perceived. Visual inspection of Q-Q plots indicated that residuals for each factor level were positively skewed, demonstrating that whilst most participants rarely reported feeling the emotion they had perceived (average proportion of emotions felt = .20), a small number of participants reported feeling emotion more frequently. The analysis was therefore repeated following square root transformation of raw data, and subsequent Q-Q plot inspection of transformed data indicated approximately normal residual distribution. However, analysis using raw and transformed data revealed equivalent findings, and therefore only the results using raw data have been reported.

There was no significant main effect for “emotion”, $F(1,38) = 2.020, p = .163$, or “musical experience”, $F(1,38) = 0.045, p = .834$, but there was a significant main effect for “modality”, $F(2,76) = 8.044, p = .001, \eta^2 = .175$, and a significant interaction between “modality” and “emotion”, $F(2,76) = 4.353, p = .016, \eta^2 = .103$ (Figure 4). Pairwise comparisons with Bonferroni correction indicated that, for angry clips, the proportion of trials where emotion was felt was lower in the auditory-only clips (.13) than in the visual-only (.19; $p = .033$; 95% CI [-0.127, -0.004]) and audiovisual clips (.25; $p < .001$; 95% CI [-0.190, -0.060]), but there was no significant difference between the visual-only and audiovisual clips ($p = .112$). For happy clips, the proportion of trials where emotion was felt was lower in the

visual-only clips (.18) than in the audiovisual clips (.26; $p = .040$; 95% CI [-0.141, -0.003]), but there were no differences between the visual-only and auditory-only clips (.22; $p = .891$), or the auditory-only and audiovisual clips ($p = .714$). Finally, within the auditory-only clips, the proportion of trials where emotion was felt was significantly higher when the expressed emotion was happy (.22) than when it was angry (.12; $p = .002$; 95% CI [0.036, 0.151]).

No significant interaction between “modality” and “musical experience” [$F(2,76) = 0.271$, $p = .763$], “emotion” and “musical experience” [$F(1,38) = 0.165$, $p = .687$], or “modality”, “musical experience”, and “emotion” [$F(2,76) = 0.595$, $p = .554$] was found.

An additional analysis was conducted on the association between years of music training and felt emotion for all the six different conditions (2 emotions: angry and happy x 3 modalities: auditory-only, visual-only and audiovisual). Years of music training did not significantly predict the level of felt emotion in the musicians’ group for any of the conditions, $F[1,18] \leq 1.620$, $p \geq 0.219$.

The same analyses carried out on the average proportion of felt emotions only for accurate responses returned the same results (see supplemental material).

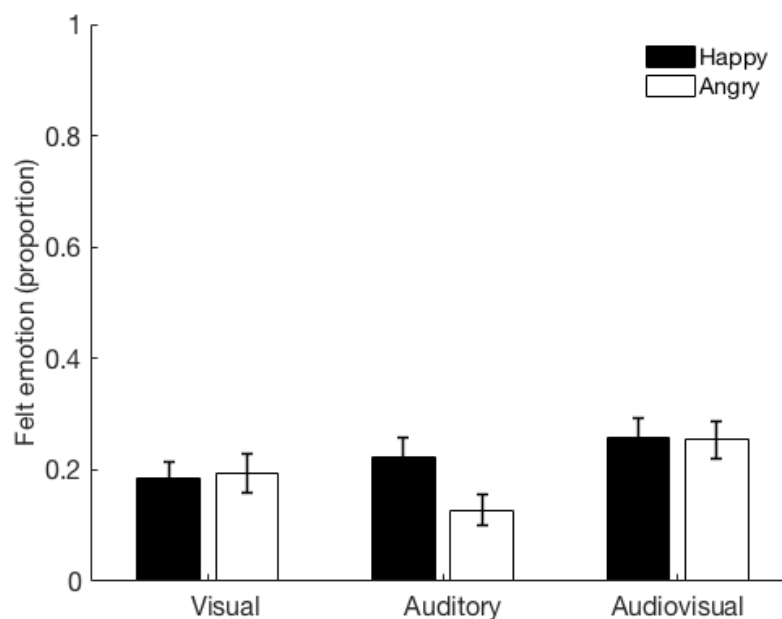


Figure 4. Proportion of happy and angry trials where participants felt the emotion they had perceived in visual-only, auditory-only and audiovisual clips, collapsing across musical experience. Error bars represent standard errors.

Since a null result does not imply the absence of a difference between musicians and non-musicians in the level of felt emotions, we also calculated Bayes Factors (BF_{10}) to evaluate the strength of evidence in favour of the null hypothesis relative to the alternate hypothesis. We used a one-way ANOVA approach with the default JZS prior (Jeffreys, 1961; Zellner and Siow, 1980; Rouder, Speckman, Sun, Morey, & Iverson, 2009) to get a Bayes Factor for each one of the six conditions (2 emotions x 3 modalities) when comparing musicians and non-musicians. We obtained for all conditions BF_{10} values ranging from .124 to .151, which are considered as moderate positive evidence in favour of the null hypothesis (e.g., Schönbrodt & Wagenmakers, 2018).

Discussion

The aim of this study was to assess whether music expertise affected the perception and feeling of emotion from others' social interaction through sound alone (speech prosody) and also through other communicative modalities such as vision (biological motion) and sound and vision (speech prosody and biological motion) combined. First, we found that musicians were more accurate than non-musicians at perceiving both happy and angry emotions within the auditory-only condition, but no differences in accuracy were found between musicians and non-musicians for the visual-only or audiovisual conditions. Second, musicians, but not non-musicians, took longer to respond to angry stimuli than to happy stimuli, but reaction times did not differ significantly between musicians and non-musicians. Finally, music training did not affect the feeling of emotion, but we found that across all

participants anger was felt significantly less often than happiness within the auditory-only condition.

Concerning emotion perception, our results are consistent with prior research by Lima and Castro (2011b), who found that musicians were more accurate than non-musicians at perceiving emotions expressed in speech. Additionally, we extend their findings in two important ways. First, we found consistent results whilst using more complex speech prosody stimuli of two people interacting, thereby indicating that the effect of music expertise on emotion recognition through speech prosody extends to multiagent social interactions. Second, by comparing performance for auditory-only, visual-only and audiovisual conditions, we were able to test whether perceptual differences between musicians and non-musicians were specific to certain communication channels. Here, our results showed that music expertise had no effect on the accuracy of emotion perception when the clips were presented in the visual-only or audiovisual conditions. This indicates that the effect of music training on emotion perception of non-musical stimuli is limited to the auditory domain, and likely occurs due to the enhanced processing of features such as pitch (Moreno et al., 2009), timbre (Chartrand & Belin, 2006), and intensity (Hausen et al., 2013) associated with extensive music training.

Moreover, we found that for non-musicians, accuracy of emotion recognition improved when stimuli from both auditory and visual modalities were presented together in the audiovisual condition, compared to the auditory-only and visual-only conditions. This result replicates the findings of Piwek et al. (2015), where the same stimuli were used, in showing that non-musicians can increase their emotion recognition accuracy through multisensory integration. Musicians, in contrast, did not show greater accuracy in the audiovisual compared to auditory-only conditions, indicating that accuracy of emotion recognition for musicians did not benefit significantly from inclusion of the visual

information. Conceptually, the integration of multisensory emotional information relies on the reliability of the sensory modalities being used (Collignon et al., 2008; Piwek et al., 2015). Indeed, multisensory facilitation (e.g. the higher accuracy for audiovisual compared to audio- and visual-only displays) is greatest when sensory cues have similar levels of reliability (Ernst & Banks, 2002), whereas facilitation diminishes when the reliability of one cue dominates the other (Alais & Burr, 2004). As such, due to enhanced processing in the auditory domain, musicians received the most reliable information through the auditory modality (even when degraded through a high-cut filter), such that the addition of visual information in the audiovisual condition did not offer any additional multisensory benefit. This partly contrasts with the findings of Vines et al. (2009) and Di Mauro et al. (2018), who found that musicians placed more weight than non-musicians on visual information when judging the expressed emotion of music clips. However, these contrasting results are arguably due to an effect of expertise and familiarity, such that musicians, who are visually familiar with music performances, are consequently more able to integrate visual information into their judgements when observing such performances.

The absence of a multisensory advantage shown by musicians in our study thus probably depends on the lack of a difference in expertise between musicians and non-musicians when judging emotions through non-musical gestures and movements. This is in agreement with previous findings by Lee and Noppeney (2011), who showed a multisensory advantage in musicians for a music clip but not for a speech clip. In fact, studies showing a multisensory advantage for musicians compared to non-musicians (Lee & Noppeney, 2011, 2014; Petrini et al., 2009; Petrini, Holt, et al., 2010; Petrini, Pollick, et al., 2011; Proverbio et al., 2015) have typically assessed musicians and non-musicians in their multisensory ability for clips representing the movements and sound of a musician. This highlights that additional

factors, such as motor experience and visual familiarity with the stimulus are likely to impact upon multisensory ability.

Across all participants, we found that happiness was identified more accurately than anger. This finding stands in contrast to numerous studies showing that participants identify angry expressions more accurately than happy expressions when listening to voices (Banse & Scherer, 1996), watching single-actor body movement (Pollick et al., 2001), and watching two people interacting (Clarke et al., 2005). Additionally, studies have also found that detection of anger elicits activation in brain regions associated with autonomic defensive behaviour in relation to threat (such as the amygdala and hypothalamus) and may therefore signify an evolutionary benefit of anger detection (Pichon, de Gelder, & Grezes, 2008). Conversely, other studies have found more comparable results to ours. For example, Belin, Fillion-Bilodeau, and Gosselin (2008) showed that non-verbal affect bursts of happiness were better recognised than anger, and Dittrich et al. (1996) found that happy displays of point-light dancers were identified more accurately than angry displays. Furthermore, using the same stimuli as used in our current study, Piwek et al. (2015) also found that happiness was identified more accurately than anger.

One explanation for the more accurate recognition of happiness over anger found in this study concerns our use of multiagent interactions. Indeed, research comparing single and multiagent stimuli has shown that participants express a bias toward perceiving emotions as angry when listening to single-speaker dialogues (Cauldwell, 2000), and demonstrate impaired recognition of happiness when viewing single-actor point-light displays (Clarke et al., 2005). This bias may be explained through consideration of the social expression of emotion. Namely, happiness is most commonly expressed through enthusiastic and dynamic social interaction involving multiple people (Shaver, Schwartz, Kirson, & O'Connor, 1987), to the extent that our perception of happiness may be impaired when the social context is

removed (Clarke et al., 2005). In contrast, anger is less often observed in public, and physical and verbal expressions of anger can frequently occur in the absence of a second person (Shaver et al., 1987). Thus, whereas the recognition of happiness in single-agent stimuli may be impaired, the recognition of anger would not be affected. Because we used multiagent stimuli, it is plausible that the perceptual bias against perceiving happiness was reduced in comparison to previous studies using single-agent displays, thus accounting for the more accurate recognition of happiness in our results. As such, our findings add further evidence toward the importance of considering multiagent social context and first- vs third-person perspective in research on emotion recognition.

Concerning reaction times, musicians, but not non-musicians, gave slower responses when recognising emotion in angry clips than in happy clips. However, reaction times did not differ significantly between musicians and non-musicians overall. A plausible explanation for this finding concerns a more pronounced speed-accuracy trade-off exhibited by musicians. Specifically, musicians are trained to carefully analyse sound, and consequently exhibit longer processing times when responding to auditory stimuli that are ambiguous or difficult to interpret (Chartrand & Belin, 2006; Münzer, Berti, & Pechmann, 2002). Musicians must also learn to carefully interpret visual cues of emotion and intention when performing or rehearsing (Gardiner, 2008); for example, to recognise when another performer will begin a solo, or to ensure they play the final note of a song in synchrony with other instrumentalists (Palmer, 1997). As such, musicians may exhibit comparable speed-accuracy trade-offs when processing both auditory and visual stimuli. In our current study, recognition of anger was less accurate than recognition of happiness, therefore indicating that angry clips were overall harder to interpret. Additionally, numerous studies have consistently indicated that happiness is the most common emotion expressed through music (Juslin & Laukka, 2004; Lindström et al., 2003), and it is therefore plausible that musicians are less familiar with expressions of

anger compared to happiness. In this case, the existence of a more pronounced speed-accuracy trade-off would explain why musicians in this current study spent more time processing the angry clips across all sensory conditions. However, we found that with the increase in years of music training there was an increase in reaction time for all the emotions and modalities, such that musicians with more training or experience took longer to recognise both happy and angry emotions in all sensory conditions (auditory, visual and audiovisual). This speed accuracy trade-off effect did not result however in higher levels of accuracy or higher levels of felt emotion in more experienced musicians, therefore indicating that the slower answers of musicians for the angry clips cannot be fully explained by a speed-accuracy trade-off.

Regarding the proportion of trials where emotions were felt as well as perceived, no differences were found between musicians and non-musicians, indicating that neither music training or accuracy of emotion recognition affected the felt emotion. We did, however, find that across all participants anger was felt significantly less often than happiness within the auditory-only condition. Here, it is possible that the expressions of anger in our study were perceived with lower intensity or clarity compared to the expressions of happiness and were consequently less likely to induce emotion. However, although this is consistent with our finding that angry expressions were recognised less accurately than happy expressions, it would not explain why the differences in felt emotion were only found for the auditory-only stimuli.

An alternative explanation concerns the importance of contextual factors in the induction of happiness and anger, such that auditory expressions of anger may induce different emotions to visual expressions of anger. Specifically, Deng and Hu (2018) and Seibt, Mühlberger, Likowski, and Weyers (2015) have shown that the induction of anger is dependent on appraisal of the situation. Indeed, visual detection of anger often signifies a

direct threat in the proximate environment and may consequently induce anger as an adaptive response (Pichon et al., 2008). Conversely, perception of anger through sound alone can indicate the presence of a threat that cannot be seen, and may therefore be more likely to induce escape-related emotions such as fear (Danesh, 1977). This is consistent with studies showing that regions associated with our own feelings of anger, such as the amygdala (Phan, Wager, Taylor, & Liberzon, 2002), are more responsive to visual than auditory expressions of anger (Costafreda, Brammer, David, & Fu, 2008). In contrast, happiness is often felt automatically in response to both visual (Deng & Hu, 2018; Seibt et al., 2015) and auditory (Johnstone, van Reekum, Oakes, & Davidson, 2006) expressions of happiness, and is less dependent on appraisal of the situation (Deng & Hu, 2018; Seibt et al., 2015). In this context, our study provides further evidence for the importance of contextual factors in the induction of emotion, such that, whilst happiness may be felt in response to both auditory and visual stimuli, anger is predominantly felt in response to visual stimuli.

Finally, the lack of an effect of music expertise on felt emotions, in contrast to the found effect of music expertise on perceived emotion supports the hypothesis that speech-music interaction might occur at an early level of processing. This agrees with reports (Musacchia, Sams, Skoe, & Kraus, 2007; Strait, Kraus, Skoe & Ashley, 2009) of musical expertise affecting subcortical (brainstem) level processing of vocal signs of emotion. The difference between perceived and felt emotion also supports the previously postulated separation between these processes in music (Gabrielsson, 2001) and extends this differentiation to voice prosody and musical expertise. Finally, although studies comparing perceived and felt/evoked emotions and differences in brain responses between musicians and non-musicians are rare, a recent study by Brattico et al. (2016) has shown a neural dissociation between perceiving sadness and happiness and evaluating pleasure-related processes from music, with musicians showing increased activation in areas related to

proprioception and salience detection, such as the insula and the anterior cingulate cortex.

Future studies could examine whether a neural dissociation between perceived and felt music is present also for vocal emotional signals and whether areas involved in processing sound are specifically modulated by musicianship in the brain network subtending emotion perception.

Limitations

Many other participants' characteristics and predispositions besides music practice and training (e.g., socio-educational background, general intellectual level or personality characteristics; Lima and Castro, 2011b) could predict or be associated to music training and thus be different in musicians and non-musicians. Our participants were mostly University students and thus the sample was quite homogeneous, and we do replicate findings from studies that did control for some of these characteristics (e.g., Lima and Castro, 2011b). However, we did not measure at what age musicians started the training, the frequency of their practice, controlled for cognitive abilities or asked for socio-educational characteristics, and thus we cannot exclude that these factors may be linked to music training and consequently to the effect of music training on emotion recognition (Strait et al., 2009; Herholz & Zatorre, 2012; Swaminathan & Schellenberg, 2018). Future studies should include these measures and use them as covariates (if not of interest) to control for their influence.

The second limitation concerns the two emotions used in our study. Happiness and anger were selected as they are both easy for actors to convey in social interactions, and both are high intensity emotions, thereby making discrimination between them more difficult (Dittrich et al., 1996; Juslin & Laukka, 2003). However, without multiple emotions to choose from, participants had to engage with a forced-choice scenario where biases in emotion misidentification may have become more problematic. Pertaining to our present study, aforementioned research has indicated that anger is misidentified as happiness more

frequently than happiness is misidentified as anger (Dittrich et al., 1996; Lima & Castro, 2011b), and it is therefore possible that participants in this current study may have been biased towards interpreting stimuli as being happy. Additionally, we cannot disentangle the effects of the specific emotions used here and emotional valence effects, as we had one positive and one negative emotion. Thus, future research would benefit from including additional emotions (both positive and negative) to avoid these types of biases and disentangle the contribution of specific emotions from that of emotional valence.

Finally, we did not ask participants to rate the intensity of the emotion they perceived in each display. Although we found interesting results based on the measures we had used, inclusion of intensity ratings would have enabled more thorough interpretation of our findings. Most notably, this would have helped determine the factors that influence felt emotions. That is, despite anger and happiness being considered high intensity emotions, in the absence of intensity ratings, it remains unclear whether, for example, the lack of difference between musicians and non-musicians in felt emotion was due to low emotional intensity portrayed by the chosen stimuli. Future studies should obtain measure of emotional intensity to clarify this point.

Conclusion

In conclusion, this current study provides empirical evidence that, through improving auditory capabilities, music training enhances perception and recognition of emotion from others' social interaction when delivered through sound, but not through vision. This indicates that the overlap between music and language processes is specific to sound, and does not generalise to other communicative channels. Furthermore, we show that music expertise enhances the ability to recognise emotions from others, but not the feeling of such emotions, indicating that the effect of music training is confined to perceptual processes of emotion.

Although the literature on the effect of music training on emotion perception from stimuli other than music still remains in its infancy, the current findings have promising implications not only theoretically but also in terms of applications. Indeed, if emotion recognition in music and speech share common auditory mechanisms that are affected by music training, music may present a useful rehabilitative device for assisting persons who have difficulty interpreting (rather than feeling) emotions from vocal expression; a symptom present in disorders such as autism (Korpilahti et al., 2007).

References

- Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3), 257-262. doi:10.1016/j.cub.2004.01.029
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614-636. doi:10.1037//0022-3514.70.3.614
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, 40(2), 531-539. doi:10.3758/BRM.40.2.531
- Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 921.
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *Journal of the Acoustical Society of America*, 137, 378–387.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, 10(4), 433-436.
- Brattico, E., Bogert, B., Alluri, V., Tervaniemi, M., Eerola, T., Jacobsen, T. (2016). It's sadbut I like it: the neural dissociation between musical emotions and liking in experts and laypersons. *Front. Hum. Neurosci.*, 9, 676.
<http://dx.doi.org/10.3389/fnhum.2015.00676>
- Castro, S. L., & Lima, C. F. (2014). Age and musical expertise influence emotion recognition in music. *Music Perception: An Interdisciplinary Journal*, 32(2), 125-142.
- Cauldwell, R. T. (2000). *Where did the anger go? The role of context in interpreting emotion in speech*. Paper presented at the ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.

- Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, 405(3), 164-167. doi:10.1016/j.neulet.2006.06.053
- Clarke, T. J., Bradshaw, M. F., Field, D. T., Hampson, S. E., & Rose, D. (2005). The Perception of Emotion from Body Movement in Point-Light Displays of Interpersonal Dialogue. *Perception*, 34(10), 1171-1180. doi:10.1068/p5203
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, 1242, 126-135. doi:10.1016/j.brainres.2008.04.023
- Costafreda, S. G., Brammer, M. J., David, A. S., & Fu, C. H. Y. (2008). Predictors of amygdala activation during the processing of emotional stimuli: A meta-analysis of 385 PET and fMRI studies. *Brain Research Reviews*, 58(1), 57-70. doi:10.1016/j.brainresrev.2007.10.012
- Dahl, S., & Friberg, A. (2007). Visual Perception of Expressiveness in Musicians' Body Movements. *Music Perception: An Interdisciplinary Journal*, 24(5), 433.
- Danesh, H. B. (1977). Anger and fear. *American Journal of Psychiatry*, 134(10), 1109-1112. doi:10.1176/ajp.134.10.1109
- Deng, H., & Hu, P. (2018). Matching Your Face or Appraising the Situation: Two Paths to Emotional Contagion. *Frontiers in Psychology*, 8(2278). doi:10.3389/fpsyg.2017.02278
- Di Mauro, M., Toffalini, E., Grassi, M., & Petrini, K. (2018). Effect of Long-Term Music Training on Emotion Perception From Drumming Improvisation. *Front. Psychol.*, 09. doi.org/10.3389/fpsyg.2018.02168
- Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of Emotion from Dynamic Point-Light Displays Represented in Dance. *Perception*, 25(6), 727-738. doi:10.1068/p250727

- Edward, P. A., & Carole, S. H. (1990). Characteristics of Motivation for Music and Musical Aptitude of Undergraduate Nonmusic Majors. *Journal of Research in Music Education, 38*(4), 258-268. doi:10.2307/3345223
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3-4), 169-200. doi:10.1080/02699939208411068
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429. doi:10.1038/415429a
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.
- Gabrielsson, A. (2002). Perceived emotion and felt emotion: same or different? *Musicae Scientiae, Special Issue 2001–2002*, 123–147.
- Gardiner, M. F. (2008). Responses to music: Emotional signaling, and learning. *Behavioral and Brain Sciences, 31*(5), 580-581. doi:10.1017/S0140525X08005359
- Good, A., Gordon, K. A., Papsin, B. C., Nespoli, G., Hopyan, T., Peretz, I., & Russo, F. A. (2017). Benefits of music training for perception of emotional speech prosody in deaf children with cochlear implants. *Ear and Hearing, 38*, 455–464. <https://doi.org/10.1097/AUD.0000000000000402>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional Contagion. *Current Directions in Psychological Science, 2*(3), 96-100. doi:10.1111/1467-8721.ep10770953
- Hausen, M., Torppa, R., Salmela, V., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: a common rhythm. *Frontiers in Psychology, 4*(566). doi:10.3389/fpsyg.2013.00566

- Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nature Neuroscience*, 6, 663. doi:10.1038/nm1080
- Herholz, Sibylle C., & Zatorre, Robert J. (2012). Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, 76(3), 486-502. doi:10.1016/j.neuron.2012.10.011
- Hill, H., Jinno, Y., & Johnston, A. (2003). Comparing Solid-Body with Point-Light Animations. *Perception*, 32(5), 561-566. doi:10.1068/p3435
- Jeffreys, H. (1961). *Theory of probability (3rd Ed.)*. Oxford, UK: Oxford University Press.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2), 201-211. doi:10.3758/BF03212378
- Johnstone, T., van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, 1(3), 242-249. doi:10.1093/scan/nsi027
- Juslin, P. N., Harmat, L., & Eerola, T. (2014). What makes music emotionally significant? Exploring the underlying mechanisms. *Psychology of Music*, 42(4), 599-623. doi:10.1177/0305735613484548
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770-814. doi:10.1037/0033-2909.129.5.770
- Juslin, P. N., & Laukka, P. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*, 33(3), 217-238. doi:10.1080/0929821042000317813
- Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *J Basic Clin Physiol Pharmacol*, 12(2 Suppl), 125-143. doi:10.1515/JBCPP.2001.12.2.125

- Klaus, R. S., Rainer, B., & Harald, G. W. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76-92. doi:10.1177/0022022101032001009
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36(14), 1.
- Knoll, M. A., Uther, M., & Costall, A. (2009). Effects of low-pass filtering on the judgment of vocal affect in speech directed to infants, adults and foreigners. *Speech Communication*, 51(3), 210-216. doi:10.1016/j.specom.2008.08.001
- Korpilahti, P., Jansson-Verkasalo, E., Mattila, M.-L., Kuusikko, S., Suominen, K., Rytty, S., . . . Moilanen, I. (2007). Processing of Affective Speech Prosody is Impaired in Asperger Syndrome. *Journal of Autism and Developmental Disorders*, 37(8), 1539-1549. doi:10.1007/s10803-006-0271-2
- Lee, H., & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences*, 108(51), E1441. doi:10.1073/pnas.1115267108
- Lee, H., & Noppeney, U. (2014). Music expertise shapes audiovisual temporal integration windows for speech, sinewave speech, and music. *Frontiers in Psychology*, 5, 868. doi:10.3389/fpsyg.2014.00868
- Lima, C. F., & Castro, S. L. (2011a). Emotion recognition in music changes across the adult life span. *Cognition and Emotion*, 25(4), 585-598. doi:10.1080/02699931.2010.502449
- Lima, C. F., & Castro, S. L. (2011b). Speaking to the Trained Ear: Musical Expertise Enhances the Recognition of Emotions in Speech Prosody. *Emotion*, 11(5), 1021-1031. doi:10.1037/a0024521

- Lima, C.F., Brancatisano, O., Fancourt, A., Müllensiefen, D., Scott, S.K., Warren, J. D., Stewart, L. (2016). Impaired socio-emotional processing in a developmental music disorder. *Scientific Reports*, 6 <https://doi.org/10.1038/srep34911>.
- Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). “Expressivity comes from within your soul”: A questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education*, 20(1), 23-47.
doi:10.1177/1321103X030200010201
- Livingstone, S. R., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing Musical Emotion: A Computational Rule System for Modifying Score and Performance. *Computer Music Journal*, 34(1), 41-64. doi:10.1162/comj.2010.34.1.41
- Lu, Y., Paraskevopoulos, E., Herholz, S. C., Kuchenbuch, A., & Pantev, C. (2014). Temporal processing of audiovisual stimuli is enhanced in musicians: evidence from magnetoencephalography (MEG). *PloS One*, 9(3), e90686.
- Ma, Y., Paterson, H. M., & Pollick, F. E. (2006). A motion capture library for the study of identity, gender, and emotion perception from biological motion. *Behavior Research Methods*, 38(1), 134-141. doi:10.3758/BF03192758
- Madsen, S. M. K., Marschall, M., Dau, T., Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, 9, 10404. doi:10.1038/s41598-019-46728-1
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. *Cerebral Cortex*, 19(3), 712-723. doi:10.1093/cercor/bhn120
- Münzer, S., Berti, S., & Pechmann, T. (2002). Encoding of timbre, speech, and tones: Musicians vs. non-musicians. *Psychological Test and Assessment Modeling*, 44(2), 187.

- Musacchia, G., Sams, M., Skoe, E., Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA*, 104: 15894–15898.
- Neumann, R., & Strack, F. (2000). "Mood contagion": the automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, 79(2), 211.
- Palmer, C. (1997). Music Performance. *Annual Review of Psychology*, 48(1), 115-138.
doi:10.1146/annurev.psych.48.1.115
- Parbery-Clark, A., Strait, D. L., Hittner, E., & Kraus, N. (2013). Musical Training Enhances Neural Processing of Binaural Sounds. *The Journal of Neuroscience*, 33(42), 16741.
doi:10.1523/JNEUROSCI.5700-12.2013
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spat Vis*, 10(4), 437-442.
- Petrini, K., Crabbe, F., Sheridan, C., & Pollick, F. E. (2011). The Music of Your Emotions: Neural Substrates Involved in Detection of Emotional Correspondence between Auditory and Visual Music Actions. *PLoS ONE*, 6(4), e19165.
doi:10.1371/journal.pone.0019165
- Petrini, K., Dahl, S., Rocchesso, D., Waadeland, C. H., Avanzini, F., Puce, A., & Pollick, F. E. (2009). Multisensory integration of drumming actions: musical expertise affects perceived audiovisual asynchrony. *Experimental Brain Research*, 198(2), 339.
doi:10.1007/s00221-009-1817-2
- Petrini, K., Holt, S. P., & Pollick, F. (2010a). Expertise with multisensory events eliminates the effect of biological motion rotation on audiovisual synchrony perception. *Journal of Vision*, 10(5), 2-2. doi:10.1167/10.5.2

- Petrini, K., McAleer, P., & Pollick, F. (2010b). Audiovisual integration of emotional signals from music improvisation does not depend on temporal correspondence. *Brain Research, 1323*, 139-148. doi:10.1016/j.brainres.2010.02.012
- Petrini, K., Pollick, F. E., Dahl, S., McAleer, P., McKay, L., Rocchesso, D., . . . Puce, A. (2011). Action expertise reduces brain activity for audiovisual matching actions: An fMRI study with expert drummers. *NeuroImage, 56*(3), 1480-1492. doi:10.1016/j.neuroimage.2011.03.009
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *NeuroImage, 16*(2), 331-348. doi:10.1006/nimg.2002.1087
- Pichon, S., de Gelder, B., & Grezes, J. (2008). Emotional modulation of visual and motor areas by dynamic body expressions of anger. *Social Neuroscience, 3*(3-4), 199-212. doi:10.1080/17470910701394368
- Pisanski, K., Fraccaro, P. J., Tigue, C. C., O'Connor, J. J. M., Röder, S., Andrews, P. W., Fink, B., DeBruine, L. M., Jones, B. C., Feinberg, D. R. (2014). Vocal indicators of body size in men and women: a meta-analysis. *Animal Behaviour, 95*, 89-99, <http://dx.doi.org/10.1016/j.anbehav.2014.06.011>.
- Piwek, L., Petrini, K., & Pollick, F. (2016). A dyadic stimulus set of audiovisual affective displays for the study of multisensory, emotional, social interactions. *Behavior Research Methods, 48*(4), 1285-1295. doi:10.3758/s13428-015-0654-4
- Piwek, L., Pollick, F., & Petrini, K. (2015). Audiovisual integration of emotional signals from others' social interactions. *Frontiers in Psychology, 6*(611). doi:10.3389/fpsyg.2015.00611
- Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51-B61. doi:10.1016/S0010-0277(01)00147-0

- Proverbio, A. M., Attardo, L., Cozzi, M., & Zani, A. (2015). The effect of musical practice on gesture/sound pairing. *Frontiers in Psychology*, 6(376).
doi:10.3389/fpsyg.2015.00376
- Roseman, I. J., & Smith, C. A. (2001). Appraisal theory: Overview, assumptions, varieties, controversies. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal Processes in Emotion: Theory, Methods, Research* (pp. 3-19). New York, NY, US: Oxford University Press.
- Rouder, J. N., Speckman P. L., Sun D., Morey R. D., & Iverson G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, 16(2), 225–237. <http://dx.doi.org/10.3758/PBR.16.2.225>
- Scherer, K. R. (1999). Appraisal Theory. In T. Dalgleish & M. J. Power (Eds.), *Handbook of Cognition and Emotion* (pp. 637-663). New York, NY, US: John Wiley & Sons Ltd.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1), 227-256. doi:10.1016/S0167-6393(02)00084-5
- Scherer, K. R., & Tannenbaum, P. H. (1986). Emotional experiences in everyday life: A survey approach. *Motivation and Emotion*, 10(4), 295-314. doi:10.1007/BF00992106
- Seibt, B., Mühlberger, A., Likowski, K., & Weyers, P. (2015). Facial mimicry in its social setting. *Frontiers in Psychology*, 6(1122). doi:10.3389/fpsyg.2015.01122
- Sharda, M., Tuerk, C., Chowdhury, R., Jamey, K., Foster, N., Custo-Blanch, M., Tan, M., Nadig, A., Hyde, K. (2018). Music improves social communication and auditory-motor connectivity in children with autism. *Transl Psychiatry*, 8:231.
- Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: further exploration of a prototype approach. *Journal of Personality and Social Psychology*, 52(6), 1061-1086.

- Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142.
- Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical Experience Promotes Subcortical Efficiency in Processing Emotional Vocal Sounds. *Annals of the New York Academy of Sciences*, 1169(1), 209-213. doi:10.1111/j.1749-6632.2009.04864.x
- Strait, D. L., O'Connell, S., Parbery-Clark, A., & Kraus, N. (2014). Musicians' Enhanced Neural Differentiation of Speech Sounds Arises Early in Life: Developmental Evidence from Ages 3 to 30. *Cerebral Cortex*, 24(9), 2512-2521.
doi:10.1093/cercor/bht103
- Swaminathan, S., & Schellenberg, E. G. (2018). Musical competence is predicted by music training, cognitive abilities, and personality. *Scientific Reports*, 8, 9223.
<https://doi.org/10.1038/s41598-018-27571-2>
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Perceiving prosody in speech: Do music lessons help? *Emotion*, 4, 46–64.
- Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, 8(6), 838-849.
doi:10.1037/a0014080
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, 118(2), 157-170. doi:10.1016/j.cognition.2010.11.010
- Wu, C. H., & Liang, W. B. (2011). Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels. *IEEE Transactions on Affective Computing*, 2(1), 10-21. doi:10.1109/T-AFFC.2010.16

- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, 8, 547. doi:10.1038/nrn2152
- Zellner, A., & Siow, A. (1980). *Posterior odds ratios for selected regression hypotheses*. In J. M. Bernardo, M. H. Degroot, D. V. Lindeley, & A. F. M. Smith (Eds.), *Bayesian Statistics: Proceedings of the first international meeting held in Valencia (Spain)* (pp. 585–603), Valencia, Spain: University of Valencia